Inferring non-neutral regulatory change in pathways from transcriptional profiling data

Joshua G. Schraiber[1], Yulia Mostovoy[2], Tiffany Y. Hsu[2,3] Rachel B. Brem[2,*]

1 Department of Integrative Biology, Univeristy of California, Berkeley, CA, USA

2 Department of Molecular and Cellular Biology, Univeristy of California, Berkeley, CA, USA

3 Present Address: Graduate Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, USA

∗ **E-mail: Corresponding rbrem@berkeley.edu**

Abstract

An outstanding question in comparative genomics is the evolutionary importance of gene expression differences between species. Rigorous molecular-evolution methods to infer evidence for natural selection from transcriptional profiling data are at a premium in the field, and to date, phylogenetic approaches have not been well-suited to address the question in the small sets of taxa profiled in standard surveys of gene expression. To meet this challenge, we have developed a strategy to infer evolutionary histories from expression data by analyzing suites of genes of common function. In a manner conceptually similar to molecular-evolution models in which the evolutionary rates of DNA sequence at multiple loci follow a gamma distribution, we modeled expression of the genes of an *a priori*-defined pathway with rates drawn from an inverse-gamma distribution. We then developed a fitting strategy to infer the parameters of this distribution from expression measurements, and to identify gene groups whose expression patterns were consistent with evolutionary constraint or rapid evolution in particular species. Simulations confirmed the power and accuracy of our inference method. As an experimental testbed for our approach, we generated and analyzed transcriptional profiles of four *Saccharomyces* yeasts. The results revealed pathways with signatures of constrained and accelerated regulatory evolution in individual yeasts, and across the phylogeny, highlighting the prevalence of pathway-level expression change during the divergence of yeast species. We anticipate that our pathway-based phylogenetic approach will be of broad utility in the search to understand the evolutionary relevance of regulatory change.

Author Summary

Comparative transcriptomic studies routinely identify thousands of genes differentially expressed between species. The central question in the field is whether and how such regulatory changes have been the product of natural selection. Can we detect the signal of evolutionarily relevant expression divergence amid the noise of changes resulting from genetic drift? Our work develops a theory of gene expression variation among a suite of genes that function together. We derive a formalism that relates empirical observations of expression of pathway genes in divergent species to the underlying strength of natural selection on expression output. We show that fitting this type of model to simulated data accurately recapitulates the parameters used to generate the simulation. We then make experimental measurements of gene expression in a panel of single-celled eukaryotic yeast species. To these data we apply our inference method, and identify pathways with striking evidence for accelerated or constrained regulatory evolution, in particular species and across the phylogeny. Our method provides a key advance over previous approaches in that it maximizes the power of rigorous molecular-evolution analysis of regulatory variation. As such, the theory and tools we have developed will likely find broad application in the field of comparative genomics.

Introduction

Comparative studies of gene expression across species routinely detect regulatory variation at thousands of loci [1–4]. Whether and how these expression changes are of evolutionary relevance has become a central question in the field. In landmark cases, experimental dissection of model phenotypes has revealed evidence for adaptive regulatory change at individual genes [5–8]. These findings have motivated hypothesis-generating, genome-scale searches for signatures of natural selection on gene regulation. In addition to molecular-evolution analyses of regulatory sequence [9–12], phylogenetic methods have been developed to infer evidence for non-neutral evolutionary change from measurements of gene expression [4, 13, 14]. Two classic models of continuous character evolution have been used for the latter purpose: Brownian motion models, which can specify lineage-specific rates of evolution on a phylogenetic tree [15–18] and have been used to model the neutral evolution of gene expression [14, 19], and the Ornstein-Uhlenbeck model, which by describing lineage-specific forces of drift and stabilizing selection [15,20,21] can be used to test for evolutionary constraint on gene expression [4, 14]. To date, phylogenetic approaches

have had relatively modest power to infer lineage-specific rates or selective optima of gene expression levels. This limitation is due in part to the sparse species coverage typical of transcriptomic surveys, in contrast to studies of organismal traits where observations in hundreds of species can be made to maximize the power of phylogenetic inference [22–24].

As a complement to model-based phylogenetic methods, more empirical approaches have also been proposed that detect expression patterns suggestive of non-neutral evolution [25–27]. We previously developed a paradigm to detect species changes in selective pressure on the regulation of a pathway, or suite of genes of common function, in the case where multiple independent variants drive expression of pathway genes in the same direction [26, 28]. Broadly, pathway-level analyses have the potential to uncover evidence for changes in selective pressure on a gene group in the aggregate, when the signal at any one gene may be too weak to emerge from genome-scale scans. However, the currently available tests for directional regulatory evolution are not well suited to cases in which some components of a pathway are activated, and others are downregulated, in response to selection.

In this work, we set out to combine the rigor of phylogenetic methods to reconstruct histories of continuous-character evolution with the power of pathway-level analyses of regulatory change. We reasoned that an integration of these two families of methods could be used to detect cases of pathway regulatory evolution from gene expression data, without assuming a polygenic or directional model. To this end, we aimed to develop a phylogenetic model of pathway regulatory change that accounted for differences in evolutionary rate between the individual genes of a pathway. We sought to use this model to uncover gene groups whose regulation has undergone accelerated evolution or been subject to evolutionary constraint, over and above the degree expected by drift during species divergence as estimated from genome sequence. As an experimental testbed for our inference strategy, we used the *Saccharomyces* yeasts. These microbial eukaryotes span an estimated 20 million years of divergence and have available well-established orthologous gene calls [29], and yeast pathways are well-annotated based on decades of characterization of the model organism *S. cerevisiae*. We generated a comparative transcriptomic data set across Saccharomycetes by RNA-seq, and we used the data to search for cases of pathway regulatory change.

Results

Modeling the rates of regulatory evolution across the genes of a pathway

The Brownian motion model of expression of a gene predicts a multivariate normal distribution of observed expression levels in the species at the tips of a phylogenetic tree. The variance-covariance matrix of this multivariate normal distribution reflects both the relatedness of the species and the rate of regulatory evolution along each branch of the tree. We sought to apply this model to interpret expression changes in a pre-defined set of genes of common function, which we term a pathway. Our goal was to test for accelerated or constrained regulatory variation in a pathway relative to the expectation from DNA sequence divergence, as specified by a genome tree. To avoid the potential for over-parameterization if the rate of each gene in a pathway were fit separately, we instead developed a formalism, detailed in Methods, to model regulatory evolution using a parametric distribution of evolutionary rates across the genes. This strategy parallels well-established models of the rate of DNA sequence evolution across different sites in a locus or genome [30]. Briefly, we assumed that each gene draws its rate of evolution in the Brownian motion model from an inverse-gamma distribution, and we derived the relationship between the parameters of this distribution and the likelihood of expression observations at the tips of the tree. This formalism enabled a maximum-likelihood fit of the distribution parameters given empirical expression data, and could accommodate models of lineage-specific regulatory evolution, in which a particular subtree was described by distinct evolutionary rate parameters relative to the rest of the phylogeny. As a point of comparison, we additionally made use of an Ornstein-Uhlenbeck (OU) model [21]: here the rate of regulatory evolution of each gene in a pathway, across the entire phylogeny, was drawn from an inverse-gamma distribution, and all genes of the pathway were subject to the same degree of stabilizing selection, again across the entire tree.

Our ultimate application of the method was to enumerate all possible Brownian motion models in which pathway expression evolved at a distinct rate along the lineages of a subtree relative to the rest of the phylogeny, and for each such model, apply our fitting strategy and tabulate the likelihood of the data under the best-fit parameter set. To compare these likelihoods and the analogous likelihood from the best-fit OU model of universal constraint, we applied a standard Akaike information criterion (AIC) [23, 31, 32] to identify strongly supported models.

Simulation testing of inference of pathway regulatory evolution

As an initial test of our approach, we sought to assess the performance of our phylogenetic inference scheme in the ideal case in which rates of regulatory evolution of the genes of a pathway were simulated from, and thus conformed to, the models of our theoretical treatment. In keeping with our experimental application below which used a comparison of *Saccharomyces* yeast species as a testbed, we developed a simulation scheme using a molecular clock-calibrated *Saccharomyces* phylogeny [29] (see Figures 1, 2, and 5). We first simulated the expression of a single gene subject to accelerated or constrained evolution in a subtree of the phylogeny. As expected, fitting a single-gene Brownian motion or OU model to these simulated data did not achieve high power or recapitulate model parameters (Figures 1-3, leftmost data points of each panel), reflecting the challenges of the phylogenetic approach when applied on a gene-by-gene basis to small trees like the *Saccharomyces* species set.

We next simulated multi-gene pathways in which the rates of regulatory evolution across the genes were drawn from an inverse-gamma distribution, with the parameters of this distribution specified to be distinct for the lineages of a given subtree relative to the rest of the phylogeny. We also simulated pathways from our OU model of a single, phylogeny-wide inverse-gamma distribution for evolutionary rates across genes of a pathway and a single, phylogeny-wide degree of constraint. And we simulated from an equal-rates model in which the rate of regulatory evolution along all branches of the tree in each gene of a pathway was drawn from the same inverse-gamma distribution. With the simulated expression data in hand from a given generating model, we fit an OU model, an equal-rates model, and models of evolutionary rate shifts in each subtree in turn. Comparing AIC weights of the likelihoods, we observed strong AIC support for the true generating model in cases of lineage-specific regulatory evolution, approaching AIC weights of 100% for the correct model in large pathways (Figure 1). In these simulations our method also inferred the correct magnitudes of lineage-specific shifts, for all but the smallest pathways (Figure 2). Likewise, when applied to simulated expression data generated under models of phylogeny-wide constraint, our method successfully identified OU as the correct model (Figure 3a), though with biased estimates of the magnitude of the constraint parameter (Figure 3b) likely due to a lack of identifiability with the inverse gamma rate parameter (Supplementary Figure 1). For both classes of model, the performance of our inference method was similar across a range of parameter values (data not shown). We conclude that our pathway-based phylogenetic approach is highly powered to infer evolutionary histories of gene expression change, particularly lineage-specific evolutionary rate shifts. As

a contrast to the poor performance of phylogenetic inference when applied to single genes, our results underscore the utility of the multi-gene paradigm in identifying candidate cases of non-neutral regulatory evolution.

Phylogenetic inference of gene expression evolution in *Saccharomyces* yeasts

We next set out to apply our method for evolutionary reconstruction of regulatory change to experimental measurements of gene expression. The total difference in gene expression between any two species is a consequence of heritable differences that act in *cis* on the DNA strand of a gene whose expression is measured, and of variants that act in *trans*, through a soluble factor, to impact gene expression of distal targets. Effects of *cis*-acting variation can be surveyed on a genomic scale using our previously reported strategy of mapping of RNA-seq reads to the individual alleles of a given gene in a diploid inter-specific hybrid [26], whereas the joint effects of *cis* and *trans*-acting factors can be assessed with standard transcriptional profiling approaches in cultures of purebred species. To apply these experimental paradigms we chose a system of *Saccharomyces sensu stricto* yeasts. We cultured two biological replicates for each of a series of hybrids formed by the mating of *S. cerevisiae* to *S. paradoxus*, *S. mikatae*, and *S. bayanus* in turn, as well as homozygotes of each species. We measured total expression in the species homozygotes, and allele-specific expression in the hybrids, of each gene by RNA-seq, using established mapping and normalization procedures (see Methods). In each set of expression data, we made use of *S. cerevisiae* as a reference: we normalized expression in the homozygote of a given species, and expression of the allele of a given species in a diploid hybrid, relative to the analogous measurement from *S. cerevisiae*.

To search for evidence of pathway regulatory evolution in our yeast expression data, we considered as pathways the pre-defined sets of genes of common function from the Gene Ontology (GO) process categories. For the genes of each GO term, we used normalized expression measurements in yeast species and, separately, measurements of *cis*-regulatory variation from interspecific hybrids, as input into our phylogenetic analysis pipeline. Thus, for each of the two classes of expression measurements, for a given GO term we fit models of lineage-specific regulatory evolution incorporating inverse-gamma-distributed rates across genes; an analogous model with no lineage-specific divergence; and an OU model of universal constraint. The results revealed a range of inferred evolutionary models and AIC support across GO terms (Figure 4, Table 1, and Table S2, Table S3), and this complete data set served as the basis for manual inspection of biologically interesting features.

A first emergent trend was the prevalence, across many GO terms, of models of distinct regulatory evolution in the lineage to *S. paradoxus* as the best fit to expression measurements in species homozygotes (Figure 4a). We noted no such recurrent model in analyses of *cis*-regulatory variation (Figure 4b), implicating *trans*-acting variants as the likely source of the regulatory divergence in *S. paradoxus*. To validate these patterns, we applied our phylogenetic inference method to expression measurements from all genes in the genome analyzed as a single group, rather than each GO term in turn. When we used expression data from species homozygotes as input for this genome-scale analysis, our method assigned complete AIC support to a model in which the rate of evolution was 2.5 times faster on the branch leading to *S. paradoxus* (AIC weight = 1), consistent with results from individual GO terms (Figure 4). An analogous inference calculation using measurements of *cis*-regulatory variation, for all genes in the genome, yielded essentially complete support for an OU model of universal constraint (AIC weight = .99). Thus, constraint on the *cis*-acting determinants of gene expression, of roughly the same degree in all yeasts, is the general rule from which changes in selective pressure on particular functions may drive deviations in individual pathways. However, for many genes, expression in the *S. paradoxus* homozygote is distinct from that of other yeasts out of proportion to its sequence divergence, suggestive of derived, *trans*-acting regulatory variants with pleotropic effects.

Among the inferences of pathway regulatory evolution from our method, we observed many cases of evolutionary interest whose best-fitting model had strong AIC support (Figure 4). For each of 15 GO terms, *cis*-regulatory expression variation measurements yielded inference of an evolutionary model with >80% AIC weight (Table 1). Many such GO terms represented candidate cases of polygenic regulatory evolution, in which multiple independent variants, at the unlinked genes that make up a pathway, have been maintained in some yeast species in response to a lineage-specific shift in selective pressure on expression of the pathway components. For example, in replicative cell aging genes (GO term 0001302), *cis*-regulatory variation measured in interspecific hybrids supported a model of polygenic, accelerated evolution in *S. paradoxus* (Figure 5a), with some pathway components upregulated and some downregulated in the latter species relative to other yeasts. The total expression levels of cell aging genes in species homozygotes were also consistent with rapid evolution in *S. paradoxus* (Figure 5a), arguing against a model of compensation between *cis*- and *trans*-acting regulatory variation, and highlighting this pathway as a particularly compelling potential case of a lineage-specific change in selective pressure.

In other instances, expression measurements in species homozygotes alone supported models of lineage-

specific evolution, with each such pathway representing a candidate case of non-neutral evolution at *trans*-acting regulatory factors. For a total of 41 GO terms, our method inferred models with >80% AIC weight from homozygote species expression data (Table 2). In these top-scoring pathways, apart from the genome-scale trend of accelerated evolution in *S. paradoxus* homozygotes (Figure 4), we also noted other lineage-specific patterns. These included a gene set annotated in transcription (GO term 0006351), whose expression levels in *S. bayanus* were less volatile than those of other yeasts and thus supported a model of lineage-specific constraint (Figure 5b). The set of top-scoring pathways from analyses of species homozygote expression data also contined some conforming to the OU model of universal constraint, such as a set of genes annotated in transport (GO term 0006281), whose expression varied less across all species than would be expected from the genome tree (Figure 5c). Taken together, our findings indicate that evolutionary histories can be inferred with high confidence from experimental measurements of pathway gene expression. In our yeast data, many pathways exhibit expression signatures consistent with non-neutral regulatory evolution in particular lineages and across the phylogeny.

Discussion

The effort to infer evolutionary histories of gene expression change has been a central focus of modern comparative genomics. Against a backdrop of a few landmark successes [4, 14], progress in the field has largely been limited by the relatively weak power of phylogenetic methods when applied, on a gene-by-gene basis, to measurements from small sets of species. In this work, we have met this challenge with a method to infer evolutionary rates of any suite of independently measured continuous characters that can be analyzed together across species. We have derived the mathematical formalism for this model, and we have illustrated the power and accuracy of our approach in simulations. We have generated yeast transcriptional profiles that complement available data sets [33, 34] by measuring *cis*-regulatory contributions to species expression differences as well as the total variation between species. With these data, we have demonstrated that our phylogenetic inference method yields robust, interpretable candidate cases of pathway regulatory evolution from experimental measurements.

The defining feature of our phylogenetic inference method is that it gains power by jointly leveraging expression measurements of a group of genes, while avoiding a high-dimensional evolutionary model. Instead of requiring an estimate of the evolutionary rate at each gene, our strategy estimates the param-

eters of a distribution of evolutionary rates across genes. Our approach thus models expression of the individual genes of a pathway as independent draws from the same distribution. Theories of composite likelihood make clear that although this independence assumption could upwardly bias the likelihoods of our best-fit models, model choice and parameter estimates will still be correct on average [35]. We note that our pathway-level approach is not contingent on the Gaussian models of regulatory evolution we have used here, and future work will evaluate the advantages of incorporating compound Poisson process [13, 36] or more general Lévy process [37] models of gene expression.

Our pathway analysis of regulatory variation used gene sets curated through Gene Ontology, but our method can easily be applied to gene modules defined on the basis of protein or genetic interactions or coexpression. Any such module is likely to contain both activators and repressors, or other classes of gene function whose expression may be quantitatively tuned in response to selection by alleles with effects of opposite sign [38, 39]. The phylogenetic approach we have developed here is well-suited to detect these non-directional regulatory patterns, rather than relying on the coherence of up- or down-regulation of pathway genes [26–28,40–42]. Ultimately, a given case of strong signal in our pathway evolution paradigm, when the best-fit model is one of lineage-specific accelerated regulatory evolution, can be explained either as a product of relaxed purifying selection or positive selection on pathway output. Our approach thus serves as a powerful strategy to identify candidates for population-genetic [28] and empirical [41, 43] tests of the adaptive importance of pathway regulatory change. We have developed an R package, IRS (Inverse gamma Rate Shifts), to facilitate the usage of our method.

The advent of RNA-seq has enabled expression surveys across non-model species in many taxa. Maximizing the biological value of these data requires methods that evaluate expression variation in the context of sequence divergence between species. As rigorous phylogenetic interpretation of expression data becomes possible, these measurements will take their place beside genome sequences as a rich source of hypotheses, in the search for the molecular basis of evolutionary novelty.

Methods

Basic model

To develop Brownian motion and Ornstein-Uhlenbeck models of pathway gene expression, we begin with the basic established forms of these models. Throughout, we use uppercase letters to represent random

variables and matrices and lowercase letters to represent nonrandom variables. Assume that we have measured expression of the genes of a pathway in n species, and that we have a fixed, time-calibrated phylogeny describing the relationships between those species. We let $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,n})$ be the observations of the expression level of the ith gene of the pathway, in each of n species. Both the Brownian motion and Ornstein-Uhlenbeck (OU) models predict that the vector $\mathbf{X}_i$ is a draw from a multivariate normal distribution with density

$$g(\mathbf{x}; \mu, \sigma^2, \mathbf{V}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n \det(\mathbf{V})}} e^{-\frac{1}{2\sigma^2}(\mathbf{x}-\mu)' \mathbf{V}^{-1} (\mathbf{x}-\mu)} \tag{1}$$

where μ is a vector representing the mean expression value at the tips of the phylogenetic tree, $\mathbf{V}$ is a covariance matrix that is specific to the process, and σ^2 describes the rate of evolution such that $\sigma^2 V_{i,j} = \mathrm{Cov}(X_i, X_j)$ where $V_{i,j}$ is the i,jth element of $\mathbf{V}$.

If we assume that there is no branch-specific directionality to evolution, we can avoid the need to estimate μ in either the Brownian motion model or the OU model by a renormalization of the data. We first arbitrarily choose the gene expression measurements in a single species (say species 1), and define the new random vector $\mathbf{Z}_i = (Z_{i,2}, Z_{i,3}, \ldots, Z_{i,n})$ by

$$Z_{i,j} = X_{i,j} - X_{i,1}.$$

By our assumption that there is no branch-specific directionality, $\mathbb{E}(X_{i,j}) = \mathbb{E}(X_{i,1})$ so $\mathbb{E}(Z_{i,j}) = 0$ for all i and j. Because each $\mathbf{X}_i$ is multivariate normally distributed with dimension n, each $\mathbf{Z}_i$ will also be multivariate normally distributed with dimension $n-1$ and a slightly different covariance structure. Letting $\mathbf{W}$ be the covariance matrix corresponding to the $\mathbf{Z}_i$, elementary calculations taking into account variances and covariances of sums of random variables reveal that

$$W_{i-1,j-1} = \begin{cases} V_{i,i} + V_{1,1} - 2V_{i,1} & \text{if } i = j \\ V_{i,j} + V_{1,1} - V_{i,1} - V_{j,1} & \text{if } i \neq j. \end{cases}$$

Next, we wish to incorporate into the Brownian motion and OU models a scheme in which the rates of evolution of the genes of a pathway are not specified independently but instead are drawn from an

inverse-gamma distribution. The inverse-gamma distribution has density

$$h(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)} e^{-\frac{\beta}{y}}, \tag{2}$$

where $\Gamma(\cdot)$ is the gamma function and α and β are shape and scale parameters. The moments of this distribution are

$$\mathbb{E}(Y) = \frac{\beta}{\alpha - 1}$$

and

$$\mathrm{Var}(Y) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)},$$

from which it follows that the inverse-gamma distribution has no mean if $\alpha < 1$ and no variance if $\alpha < 2$. These properties allow for the distribution of rates of gene expression evolution in a pathway to be relatively broad; on the other hand, the inverse gamma density has no mass at 0, which prevents any gene in a pathway from not evolving at all. In addition, as $\alpha \to \infty$ and $\beta \to \infty$ as $\frac{\beta}{\alpha-1} = \mu$ stays fixed, the distribution converges to a point mass at μ. Thus, a model where there is one rate for every gene is nested within the inverse-gamma distributed rates model.

Computation of the density of the vector of expression measurements $\mathbf{Z}_i$ under this model is simplified by the fact that the inverse-gamma distribution is the conjugate prior to the variance of a normal distribution. Hence, we see that the density of $\mathbf{Z}$ is

$$\begin{aligned} f(\mathbf{z}) &= \int_0^\infty g(\mathbf{z}; \sigma^2, \mathbf{W}) h(\sigma^2) d(\sigma^2) \\ &= \int_0^\infty \frac{1}{\sqrt{(2\pi\sigma^2)^{n-1}\det(\mathbf{W})}} e^{-\frac{1}{2\sigma^2}\mathbf{z}'\mathbf{W}^{-1}\mathbf{z}} \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} e^{-\frac{\beta}{\sigma^2}} d(\sigma^2) \\ &= \frac{1}{\sqrt{(2\pi)^{n-1}\det(\mathbf{W})}} \frac{\beta^\alpha}{\left(\frac{1}{2}\mathbf{z}'\mathbf{W}^{-1}\mathbf{z} + \beta\right)^{\alpha+(n-1)/2}} \frac{\Gamma(\alpha + (n-1)/2)}{\Gamma(\alpha)}. \end{aligned} \tag{3}$$

Thus, the likelihood of the observations of transcriptome-wide gene expression across the pathway in

n taxa, normalized by the expression level in taxon 1, is

$$L(\alpha,\beta)=\prod_i \frac{1}{\sqrt{(2\pi)^{n-1}\det(\mathbf{W})}}\frac{\beta^\alpha}{(\frac{1}{2}(\mathbf{z}_i'\mathbf{W}^{-1}\mathbf{z}_i+\beta)^{\alpha+(n-1)/2}}\frac{\Gamma(\alpha+(n-1)/2)}{\Gamma(\alpha)}. \quad (4)$$

For the application to simulated and experimental data as described below, given observations of gene expression of the species at the tips of the tree, and a model that specifies the covariance matrix $\mathbf{V}$, we optimized the log likelihood function using the L-BFGS-B optimization routine in R [44].

Covariance matrix

In the previous section, we left the unnormalized covariance matrix $\mathbf{V}$ unspecified. Here we briefly recall the forms of $\mathbf{V}$ under Brownian motion and the Ornstein-Uhlenbeck process. Define the height of the evolutionary tree to be T and and the height of the node containing the common ancestor of taxa i and j by t_{ij}. Then the covariance matrix for Brownian motion is

$$V_{i,j}=\begin{cases} t_{ij} & \text{if } i\neq j \\ T & \text{if } i=j \end{cases}$$

and the covariance matrix for the Ornstein-Uhlenbeck process is

$$V_{i,j}=\begin{cases} \frac{1}{2\theta}e^{-2\theta(T-t_{ij})}(1-e^{2\theta t_{ij}}) & \text{if } i\neq j \\ \frac{1}{2\theta}(1-e^{2\theta T}) & \text{if } i=j \end{cases}$$

where θ quantifies the strength of stabilizing selection (large θ corresponds to stronger selection).

To model lineage-specific shifts in the evolutionary rate of gene expression in the context of the Brownian motion model, we adopt a framework similar to that of O'meara *et al.* [17]. We assume that in a specified subtree of the total phylogeny, the rate of evolution of every gene is multiplied by a constant, compared to the rest of the tree. Under the Brownian motion model, this is equivalent to multiplying the branch lengths in that part of the tree by that same constant; hence, shifts in evolutionary rate are incorporated by multiplying the appropriate elements of $\mathbf{W}$ by the value of the rate shift.

Comparing likelihoods among fitted models

To evaluate the support for the distinct models we fit to expression data for a given pathway, we require a strategy that will be broadly applicable in cases where no *a priori* expectation of the correct model is available, such that nested hypothesis testing schemes [17] are not applicable. Instead, given likelihoods L from fitting of each model in turn to expression data from the genes of a pathway, we use the Akaike Information Criterion, $2k - 2ln(L)$ [45], to report the strength of the support for each, where k is the number of parameters in the model ($k = 2$ for the Brownian motion model in which the rate of evolution is the same along all lineages in the phylogeny, and $k = 3$ for all other models).

Simulations

For all simulations, we used a phylogenetic tree adapted from Scannell *et al.* by removing the branch leading to *Saccharomyces kudriavzevii* (see insets of Figures 1, 2 and 5). To perform simulations, we generated expression data for one gene at a time as follows. We first drew the rate of evolution from the appropriately parameterized inverse-gamma distribution. Then, without loss of generality, we specified that the expression level at the root of the phylogeny was equal to 0, and we simulated evolution along the branches of the Scannell *et al.* tree according to either a Brownian motion or an Ornstein Uhlenbeck process (with optimal expression level equal to 0), using the terminal expression level on a branch as the initial expression level of its daughter branches. To account for lineage-specific shifts in evolutionary rate in a simulated pathway, we multiplied the rate of evolution of each gene by the rate shift parameter for evolution along the branches affected by the rate shift. For each Brownian motion-based rate shift model applicable to the tree (see insets of Figure 1 and 2), we simulated 100 replicate datasets for each of a range of gene group sizes, in each case setting $\alpha = 3$, $\beta = 2$, and a rate shift parameter equal to 5. For the Ornstein-Uhlenbeck model, we simulated 100 replicate datasets for each of a range of pathway sizes with $\alpha = 3$, $\beta = 2$, and $\theta = 10$.

Yeast strains, growth conditions, and RNA-seq

Strains used in this study are listed in Table S1. For pairwise comparisons of S. cerevisiae and each of *S. paradoxus*, *S. mikatae*, and *S. bayanus*, two biological replicates of each diploid parent species and each interspecific hybrid were grown at 25°C in YPD medium [46] to log phase (between 0.65-0.75

OD at 600 nm). Total RNA was isolated by the hot acid phenol method [46] and treated with Turbo DNA-free (Ambion) according to the manufacturer's instructions. Libraries for a strand-specific RNA-seq protocol on the Illumina sequencing platform, which delineates transcript boundaries by sequencing poly-adenylated transcript ends, were generated as in [47] with the following modifications: 1) AmpureXP beads (Beckman) were used to clean up enzymatic reactions; 2) the gel purification and size-selection step was eliminated; 3) the oligo-dT primer used for cDNA synthesis was phosphorothioated at position ten (TTTTTTTTTT*TTTTTTTTTTVN, V=A,C,G, N=A,C,G,T, *=phosphorothioate linkage, Integrated DNA Technologies); and 4) 12 PCR cycles were performed. Libraries were sequenced using 36 bp paired-end modules on an Illumina IIx Genome Analyzer (Elim Biopharmaceuticals).

RNA-seq mapping and normalization

Bioinformatic analyses were conducted in Python and R. RNA-seq reads were stripped of their putative poly-A tails by removing stretches of consecutive Ts flanking the sequenced fragment; reads without at least two such Ts were discarded, as were reads with Ts at both ends. To ensure that expression data from hybrid diploids and purebred species could be compared, for each class of expression measurement for a given pair of species we mapped reads to both species genomes from http://www.saccharomycessensustricto.org [29] using Bowtie [48] with default settings and flags -m1 -X1000. These settings allowed us to retain only those reads that were unambiguously assigned to one of the two species in each pairwise comparison. A mapped read was inferred to have originated from the plus strand of the genome if its poly-A tail corresponded to a stretch of As at the $3'$ end of the fragment, and a read was assigned to the minus strand if its poly-A tail corresponded to a stretch of Ts at the $5'$ end of the fragment relative to the reference genome. To filter out cases in which inferred poly-A tails originated from stretches of As or Ts encoded endogenously in the genome, we eliminated from analysis all reads whose stretch of As or Ts contained more than 50% matches to the reference genome. In order to filter out cases of potential oligo-dT mispriming during cDNA synthesis, we also eliminated from analysis all reads that contained 10 or more As in the 20 nucleotides upstream of their transcription termination site.

We controlled for read abundance biases due to differing GC content as follows. For each lane of sequencing, we grouped sets of overlapping reads and normalized abundance according to GC content of the overlapping region using full-quantile normalization as implemented in the package EDASeq [49]. Normalized abundance was divided by raw abundance to generate a weight that was assigned to every

read in the group. These weights were used in place of raw read counts in all downstream analyses. All expression data are available through the Gene Expression Omnibus under identification number GSE38875.

Transcript annotation

Coordinates of orthologous open reading frames (ORFs) in each genome were taken from http://www.saccharomycessensust These ORF boundaries in *S. cerevisiae* differed, in some cases, from ORF definitions in the *Saccharomyces* Genome Database [50, SGD, using the definitions from December 22, 2007]; genes for which the two sets of definitions did not overlap were discarded. For cases where the definitions overlapped but differed by more than ten base pairs at either end, we used the boundaries defined by SGD and adjusted ortholog boundaries in other species accordingly after performing local multiple alignment [51] of the orthologous regions and flanking sequences as defined by [29].

For most genomic loci, each sense transcript feature was defined as the region from 50 bp upstream to 500 bp downstream of its respective ORF. If sequence within this window for a given target ORF overlapped with the boundaries of an adjacent gene or known non-coding RNA on the same strand, the sense feature boundaries of the target were trimmed to eliminate the overlap. For tandem gene pairs, the $3'$ boundary of the upstream gene sense feature was set to 500 bp past the coding stop or the coding start of the downstream gene sense feature, whichever was closer; the $5'$ boundary of the downstream gene sense feature was set to 50 bp upstream of its coding start or the $3'$ end of the upstream gene sense feature, whichever was closer.

We tabulated the GC-normalized expression counts (see above) that mapped to each transcript feature for each RNA-seq sample. Given the full set of such counts across all features and all samples, we then applied the upper-quartile between-lane normalization method implemented in EDASeq [49]. The normalized counts from this latter step for a given species were averaged across all biological replicates to yield a final expression level for the feature, used in all analysis in this work.

Yeast pathways

We downloaded the list of genes associated with each Gene Ontology process term from the *Saccharomyces* Genome Database and filtered for terms containing at least 10 genes. The resulting set comprised 333 terms.

Visualizing distributions of interspecific expression variation

For visual inspection of expression differences between species in Figure 5, we normalized experimentally measured data by branch length as follows. If expression evolution follows the same Gaussian-based model on all lineages of the yeast phylogeny, when the expression level of gene j in taxon i is compared to that in taxon 1 used as a reference, the marginal distribution $Z_{i,j}$ (the difference in expression between taxon i and taxon 1 at gene j) is distributed according to a univariate analog of equation (3). In this case, dividing $Z_{i,j}$ by the absolute branch length between taxon i and taxon 1 eliminates the dependence of the distribution on the total divergence time between taxa, and the density of this normalized quantity will be the same for all species comparisons. In the case of lineage-specific shifts in evolutionary rate or universal selective constraint, one or more taxa will exhibit distinct densities of the normalized expression divergence measure. Thus, we generated each distribution in Figure 5 by tabulating the log fold-change in expression between the indicated species and *S. cerevisiae*, and then dividing this quantity by the divergence time between the indicated species and *S. cerevisiae* according to the genome tree.

Acknowledgments

The authors thank Davide Risso and Oh Kyu Yoon for generously providing software before publication, and John Huelsenbeck, Mason Liang, Nicholas Matzke, Rasmus Nielsen, Benjamin Peter, Jeremy Roop, and Montgomery Slatkin for helpful discussions.

References

1. Tirosh I, Reikhav S, Levy AA, Barkai N (2009) A yeast hybrid provides insight into the evolution of gene expression regulation. Science 324: 659–662.

2. Emerson J, Hsieh LC, Sung HM, Wang TY, Huang CJ, et al. (2010) Natural selection on cis and trans regulation in yeasts. Genome Research 20: 826–836.

3. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, et al. (2010) Regulatory divergence in drosophila revealed by mRNA-seq. Genome Research 20: 816–825.

4. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, et al. (2011) The evolution of gene expression levels in mammalian organs. Nature 478: 343–348.

5. Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB (2009) Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population. Science 326: 1663–1667.

6. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327: 302–305.

7. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55–61.

8. Ishii T, Numaguchi K, Miura K, Yoshida K, Thanh PT, et al. (2013) OsLG1 regulates a closed panicle trait in domesticated rice. Nature Genetics .

9. Torgerson DG, Boyko AR, Hernandez RD, Indap A, Hu X, et al. (2009) Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. PLoS Genetics 5: e1000592.

10. He BZ, Holloway AK, Maerkl SJ, Kreitman M (2011) Does positive selection drive transcription factor binding site turnover? A test with Drosophila cis-regulatory modules. PLoS Genetics 7: e1002053.

11. Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, et al. (2012) Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. PLoS Genetics 8: e1002789.

12. Gronau I, Arbiza L, Mohammed J, Siepel A (2013) Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. Molecular Biology and Evolution .

13. Chaix R, Somel M, Kreil DP, Khaitovich P, Lunter GA (2008) Evolution of primate gene expression: drift and corrective sweeps? Genetics 180: 1379–1389.

14. Bedford T, Hartl DL (2009) Optimization of gene expression by natural selection. Proceedings of the National Academy of Sciences 106: 1133–1138.

15. Lande R (1976) Natural selection and random genetic drift in phenotypic evolution. Evolution 30: 314–334.

16. Felsenstein J (1988) Phylogenies and quantitative characters. Annual Review of Ecology and Systematics 19: 445–471.

17. O'Meara BC, Ané C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution using likelihood. Evolution 60: 922–933.

18. Eastman JM, Alfaro ME, Joyce P, Hipp AL, Harmon LJ (2011) A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65: 3578–3589.

19. Oakley TH, Gu Z, Abouheif E, Patel NH, Li WH (2005) Comparative methods for the analysis of gene-expression evolution: an example using yeast functional genomic data. Molecular Biology and Evolution 22: 40–50.

20. Hansen TF (1997) Stabilizing selection and the comparative analysis of adaptation. Evolution 51: 1341–1351.

21. Butler MA, King AA (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. The American Naturalist 164: 683–695.

22. Ané C (2008) Analysis of comparative data with hierarchical autocorrelation. The Annals of Applied Statistics : 1078–1102.

23. Harmon LJ, Losos JB, Jonathan Davies T, Gillespie RG, Gittleman JL, et al. (2010) Early bursts of body size and shape evolution are rare in comparative data. Evolution 64: 2385–2396.

24. Boettiger C, Coop G, Ralph P (2012) Is your phylogeny informative? Measuring the power of comparative methods. Evolution 66: 2240–2251.

25. Blekhman R, Oshlack A, Chabot AE, Smyth GK, Gilad Y (2008) Gene regulation in primates evolves under tissue-specific selection pressures. PLoS Genetics 4: e1000271.

26. Bullard JH, Mostovoy Y, Dudoit S, Brem RB (2010) Polygenic and directional regulatory evolution across pathways in Saccharomyces. Proceedings of the National Academy of Sciences 107: 5058–5063.

27. Fraser HB, Moses AM, Schadt EE (2010) Evidence for widespread adaptive evolution of gene expression in budding yeast. Proceedings of the National Academy of Sciences 107: 2977–2982.

28. Martin HC, Roop JI, Schraiber JG, Hsu TY, Brem RB (2012) Evolution of a membrane protein regulon in Saccharomyces. Molecular Biology and Evolution 29: 1747–1756.

29. Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, et al. (2011) The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the Saccharomyces sensu stricto genus. G3: Genes, Genomes, Genetics 1: 11–25.

30. Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. Trends in Ecology and Evolution 11: 367–372.

31. Gong Z, Matzke NJ, Ermentrout B, Song D, Vendetti JE, et al. (2012) Evolution of patterns on Conus shells. Proceedings of the National Academy of Sciences 109: E234–E241.

32. Slater GJ, Harmon LJ, Alfaro ME (2012) Integrating fossils with molecular phylogenies improves inference of trait evolution. Evolution 66: 3931–3944.

33. Busby MA, Gray JM, Costa AM, Stewart C, Stromberg MP, et al. (2011) Expression divergence measured by transcriptome sequencing of four yeast species. BMC Genomics 12: 635.

34. Goodman AJ, Daugharthy ER, Kim J (2013) Pervasive antisense transcription is evolutionarily conserved in budding yeast. Molecular Biology and Evolution 30: 409–421.

35. Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. Statistica Sinica 21: 5–42.

36. Khaitovich P, Pääbo S, Weiss G (2005) Toward a neutral evolutionary model of gene expression. Genetics 170: 929–939.

37. Landis MJ, Schraiber JG, Liang M (2013) Phylogenetic analysis using Lévy processes: Finding jumps in the evolution of continuous traits. Systematic Biology 62: 193–204.

38. Maughan H, Birky CW, Nicholson WL (2009) Transcriptome divergence and the loss of plasticity in Bacillus subtilis after 6,000 generations of evolution under relaxed selection for sporulation. Journal of Bacteriology 191: 428–433.

39. Howden BP, McEvoy CR, Allen DL, Chua K, Gao W, et al. (2011) Evolution of multidrug resistance during Staphylococcus aureus infection involves mutation of the essential two component regulator WalKR. PLoS Pathogens 7: e1002359.

40. Fraser HB, Babak T, Tsang J, Zhou Y, Zhang B, et al. (2011) Systematic detection of polygenic cis-regulatory evolution. PLoS genetics 7: e1002023.

41. Fraser HB, Levy S, Chavan A, Shah HB, Perez JC, et al. (2012) Polygenic cis-regulatory adaptation in the evolution of yeast pathogenicity. Genome Research 22: 1930–1939.

42. Fraser HB (2013) Gene expression drives local adaptation in humans. Genome Research .

43. Booth LN, Tuch BB, Johnson AD (2010) Intercalation of a new tier of transcription regulation into an ancient circuit. Nature 468: 959–963.

44. Zhu C, Byrd RH, Lu P, Nocedal J (1997) Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. ACM Transactions on Mathematical Software 23: 550–560.

45. Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19: 716–723.

46. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman J, et al. (2002) Short protocols in molecular biology: a compendium of methods from current protocols in molecular biology, volume 2. Wiley New York.

47. Yoon OK, Brem RB (2010) Noncanonical transcript forms in yeast and their regulation during environmental stress. RNA 16: 1256–1267.

48. Langmead B, Trapnell C, Pop M, Salzberg SL, et al. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10: R25.

49. Risso D, Schwartz K, Sherlock G, Dudoit S (2011) GC-content normalization for RNA-seq data. BMC Bioinformatics 12: 480.

50. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, et al. (1998) SGD: Saccharomyces genome database. Nucleic Acids Research 26: 73–79.

517 51. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput.
518 Nucleic Acids Research 32: 1792–1797.

Figure Legends

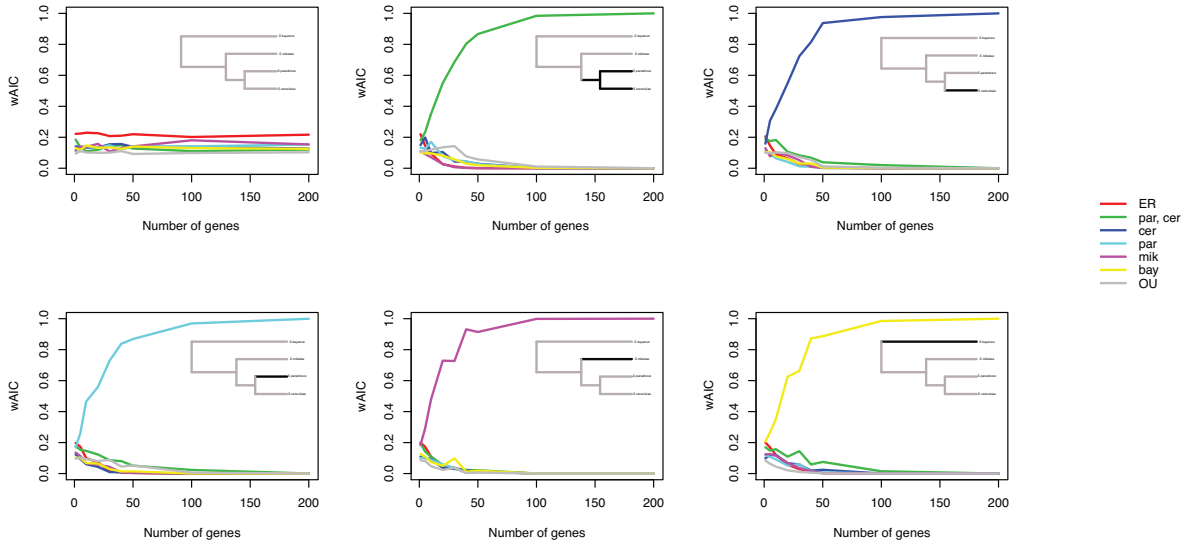


Figure 1. Phylogenetic inference of the correct model of pathway regulatory evolution from data simulated under Brownian motion models. Each panel shows results of analyses of pathway gene expression data simulated under one Brownian motion-based evolutionary model, with dark lines in each inset cartoon indicating lineages subject to accelerated regulatory evolution in the respective simulation. In a given panel, each trace reports results from maximum-likelihood fitting of the indicated model (legend at right) to simulated data from pathways of varying size; the x axis reports the number of genes in a pathway and the y axis reports the Akaike weight of the indicated model. Each data point represents results of simulations of the indicated pathway size in which the initial evolutionary rate of each pathway gene was drawn from an inverse-gamma distribution with $\alpha = 3, \beta = 2$ and then increased by a factor of 5 for lineages shown in dark grey. In the legend, ER denotes a Brownian-motion model of equal evolutionary rates on all branches of the phylogeny, OU denotes an Ornstein-Uhlenbeck model of constraint on all branches, and species names denote Brownian-motion models of increased evolutionary rate on the subtrees leading to the respective taxa.

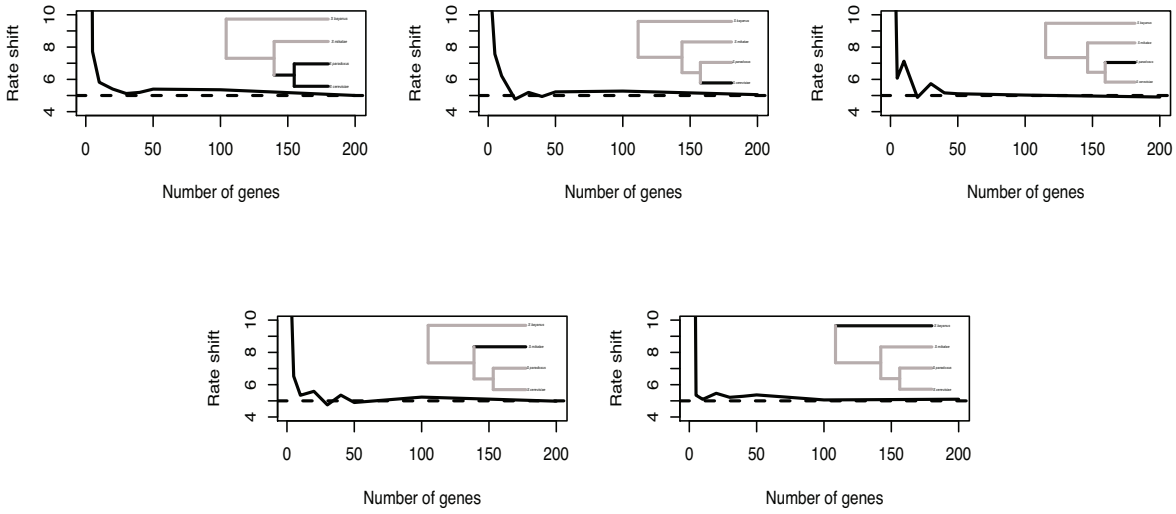


Figure 2. Phylogenetic inference of the correct degree of accelerated regulatory evolution from data simulated under Brownian motion models. Each panel shows results of analyses of pathway gene expression data simulated under one Brownian motion-based evolutionary model. Simulations and fitting, and the reporting of results, are as in Figure 1, except that in a given panel, the y axis reports the shift in the rate of regulatory evolution on the indicated lineage, relative to that on the rest of the phylogeny. The black trace reports maximum-likelihood fitted values from analysis of simulated data and the dashed line indicates the true value in the generating model of the simulation.

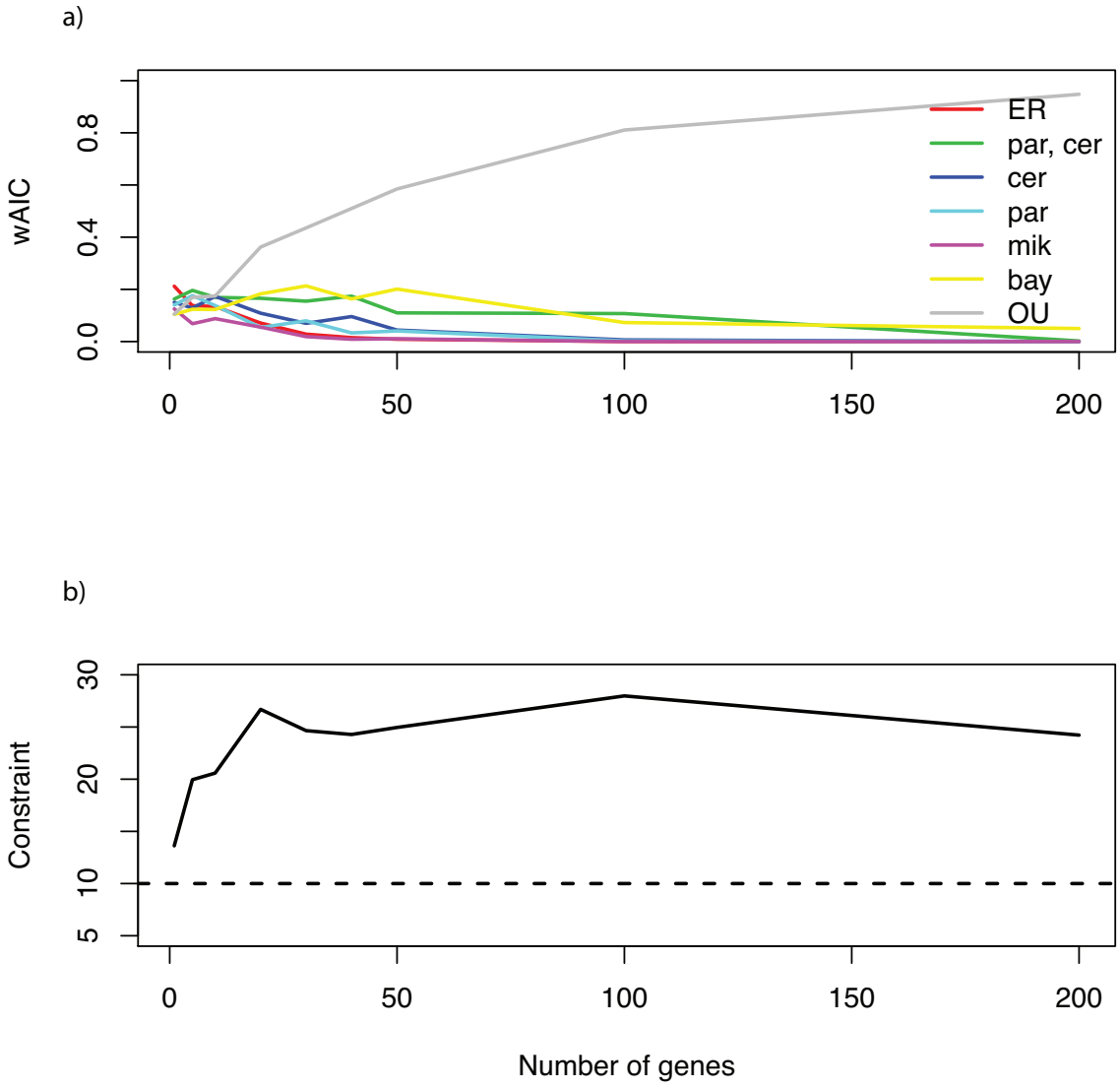


Figure 3. Phylogenetic inference of the evolutionary history of pathway regulation from data simulated under an Ornstein-Uhlenbeck model. Simulations were performed as in Figure 1 except that expression data were simulated under an Ornstein-Uhlenbeck model of universal constraint across the yeast phylogeny. (a), The x axis reports pathway size and the y axis reports Akaike weights of fitted models as in Figure 1. (b), The x axis reports pathway size and the y axis reports the values of the Ornstein-Uhlenbeck constraint parameter θ. The black trace reports maximum-likelihood fitted values from analysis of simulated data and the dashed line indicates the true value in the generating model of the simulation.

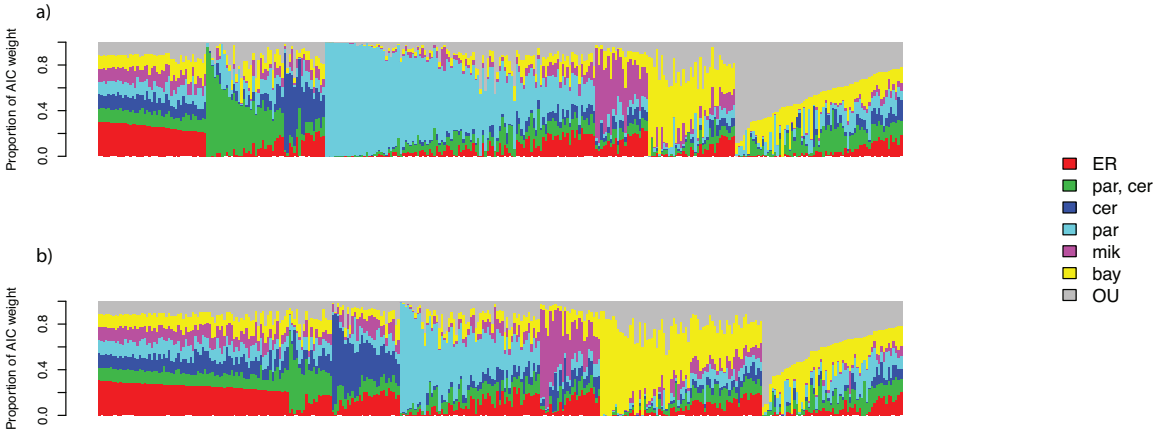


Figure 4. Inference of regulatory evolution in yeast pathways. Each panel reports results of phylogenetic inference of evolutionary histories of gene expression change from one set of experimental transcriptional profiling data. In a given panel, each vertical bar reports results of maximum-likelihood fits of Brownian-motion and Ornstein-Uhlenbeck models to expression of the genes of one Gene Ontology process term; the total proportion of a bar corresponding to a particular color indicates the Akaike weight of the corresponding model (legend at right, with labels as in Figure 1). (a), Inference from total expression measurements in homozygote species. (b), Inference from measurements of *cis*-regulatory variation in interspecific hybrids.

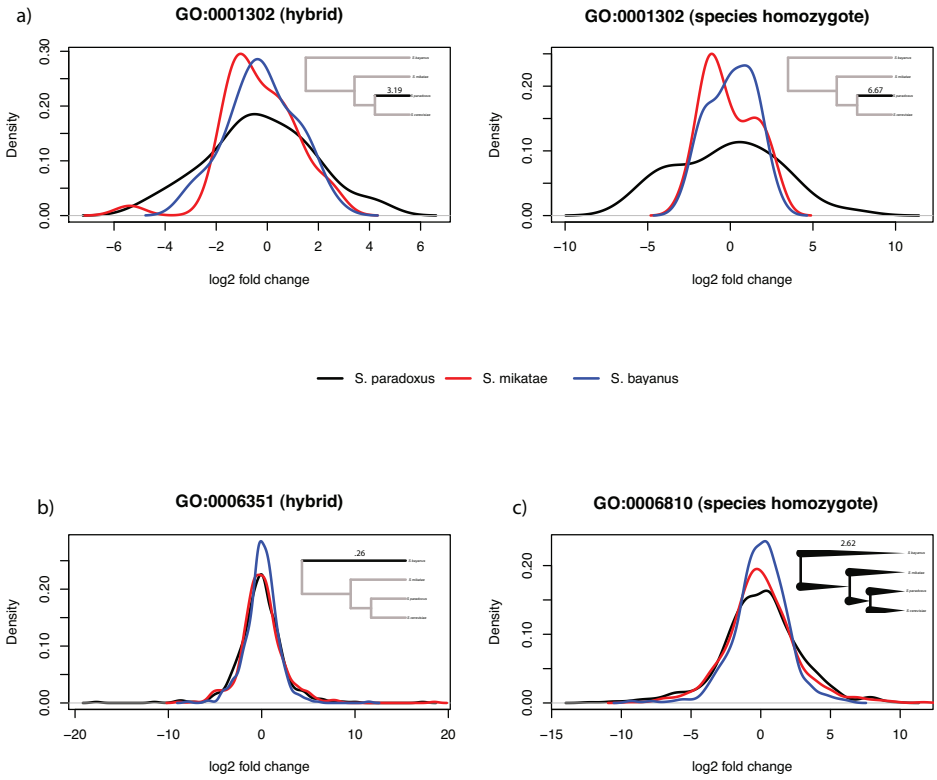


Figure 5. Lineage-specific regulatory evolution and constraint in yeast pathways. Each panel shows distributions of experimental gene expression measurements among the genes of one yeast Gene Ontology process term whose evolutionary history was inferred with strong support. In a given panel, each trace reports the expression levels of the genes of the indicated pathway, from the allele of the indicated yeast species in a hybrid or in the purebred homozygote of a species, normalized with respect to the analogous measurement in *S. cerevisiae* and with respect to branch length. Inset cartoons represent the model inferred with AIC weight >80% for the indicated pathway (see Tables 1 and 2). (a) Allele-specific expression from measurements in diploid hybrids (left) and total expression measurements in species homozygotes (right) for the genes of GO:0001302, replicative cell aging, supporting a model of accelerated evolution in *S. paradoxus*; in the inset, the number above the bolded branch reports the inferred shift in the rate of regulatory evolution along that lineage. (b) Allele-specific expression from measurements in diploid hybrids for the genes of GO:0006351, transport, supporting a model of constraint in *S. bayanus*; in the inset, the number above the bolded branch reports the inferred shift in the rate of regulatory evolution along that lineage. (c) Total expression measured in species homozygotes for the genes of GO:0006281, transcription, supporting an Ornstein-Uhlenbeck model of universal constraint; in the inset, the number above the tree reports the inferred value of the constraint parameter. Note that in (c), results are consistent with the effect of selective constraint driving expression measurements in each taxon to revert to a universal mean, such that normalizing expression measurements by total evolutionary time overcorrects the divergence of long branches.

Tables

Table 1. Top-scoring fitted models of *cis*-regulatory evolution in yeast pathways.

GO term	Model	wAIC	Constraint or shift parameter
34599	Ornstein-Uhlenbeck	0.899405768	49.97745883
6355	*S. bayanus* shift	0.837382338	0.230918849
6351	*S. bayanus* shift	0.849912647	0.258701476
1302	*S. paradoxus* shift	0.859866949	3.197059161
6897	*S. paradoxus* shift	0.965743399	4.292287639
6338	*S. cerevisiae* shift	0.840339574	0.037806902
42254	Ornstein-Uhlenbeck	0.924785133	3.733770466
6364	Ornstein-Uhlenbeck	0.902358815	3.079387696
44255	*S. paradoxus* shift	0.945799302	11.43989834
54	*S. paradoxus* shift	0.91523272	9.314688245
16310	*S. bayanus* shift	0.902247359	0.188381056
8152	*S. bayanus* shift	0.844716856	0.043114988
6629	*S. bayanus* shift	0.91650274	0.005082617
122	*S. bayanus* shift	0.819216472	0.040060263
30437	*S. paradoxus* shift	0.931136455	4.060128813

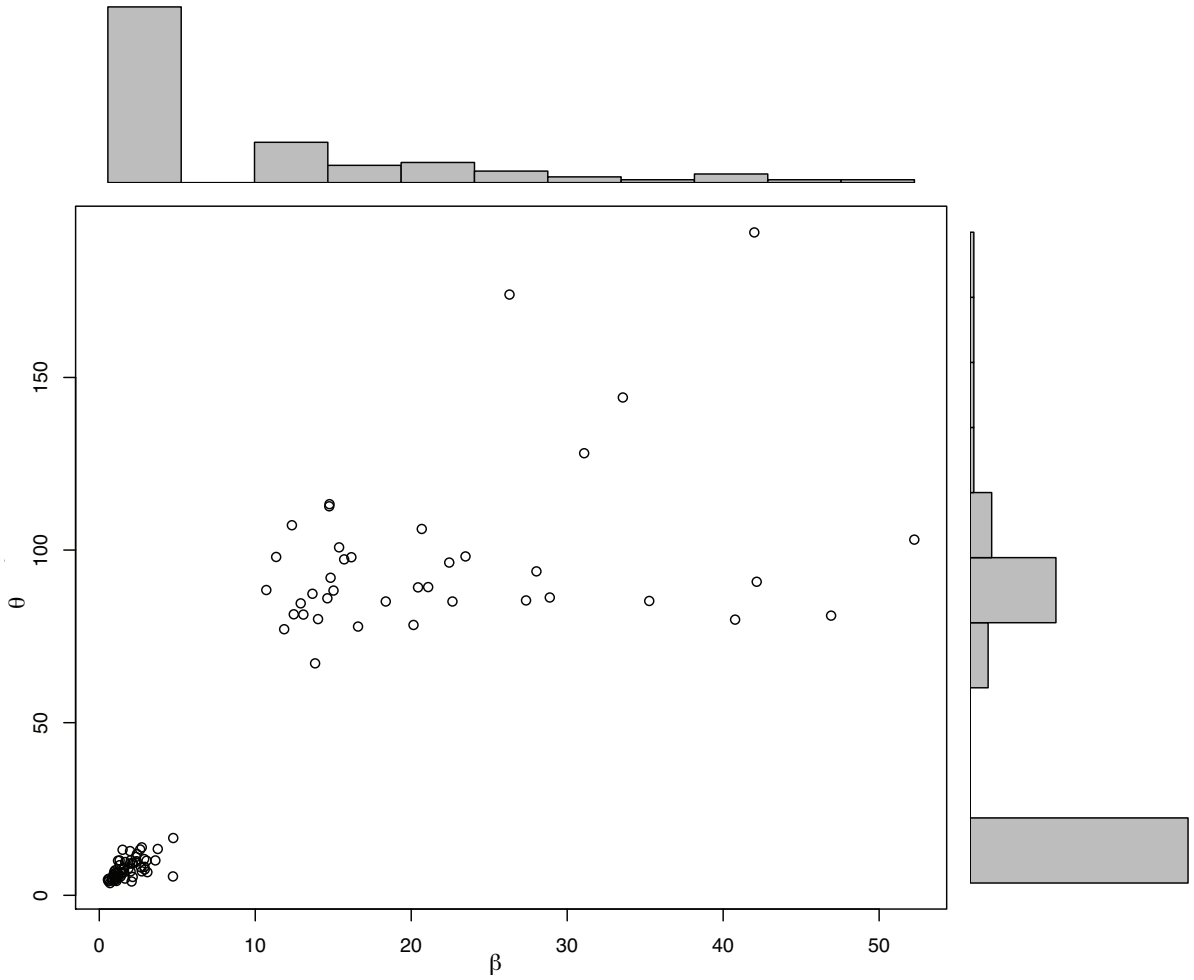
Each row reports the results of phylogenetic inference of the evolutionary history of gene regulation for one yeast Gene Ontology process term, from experimental measurements of *cis*-regulatory variation in interspecific yeast hybrids. Model, best-fit model from among the five possible Brownian motion models of evolutionary rate shift in lineages of the *Saccharomyces* phylogeny (see Figure 1), the Ornstein-Uhlenbeck (OU) model of universal constraint, and the equal-rates model involving no lineage-specific differences in evolutionary rate. wAIC, Akaike Information Criterion weight of the indicated model. Constraint or shift parameter, fitted value of the strength of purifying selection or the shift in the rate of regulatory evolution on the indicated lineage, when the best-fit model was the OU model of constraint or a Brownian motion lineage-specific evolutionary rate model, respectively.

Table 2. Top-scoring fitted models of species regulatory evolution in yeast pathways.

GO term	Model	wAIC	Constraint or shift parameter
6397	*S. paradoxus* shift	0.965171603	3.028130303
8033	*S. paradoxus* shift	0.969683391	3.714749932
71038	*S. paradoxus* shift	0.89725301	6.751073973
480	*S. paradoxus* shift	0.928296518	4.460579672
42274	*S. paradoxus* shift	0.958076119	8.083546161
472	*S. paradoxus* shift	0.953733629	4.686648741
15031	*S. bayanus* shift	0.872939854	0.183834463
1302	*S. paradoxus* shift	0.999927135	6.671016575
6006	*S. paradoxus* shift	0.816341854	4.6555377
6260	*S. paradoxus* shift	0.831407464	3.043207869
30163	*S. paradoxus* shift	0.82364567	7.009201233
6897	*S. paradoxus* shift	0.970677101	4.408614609
6412	*S. paradoxus* shift	0.981277345	2.770778823
7121	*S. paradoxus* shift	0.998579562	16.81960721
6914	Ornstein-Uhlenbeck	0.810293525	41.38598192
30488	*S. paradoxus* shift	0.893282646	7.945094861
42254	*S. paradoxus* shift	0.99999983	6.856141937
6200	*S. paradoxus* shift	0.81144199	5.590943868
6468	*S. paradoxus* shift	0.990399439	2.655209273
16567	*S. paradoxus* shift	0.959694914	3.313920599
6364	*S. paradoxus* shift	0.999995709	5.841035759
6754	*S. paradoxus* shift	0.816303046	4.668929462
422	Ornstein-Uhlenbeck	0.877576591	57.08946364
463	*S. paradoxus* and *S. cerevisiae* shift	0.958484282	10.39289039
6414	*S. paradoxus* and *S. cerevisiae* shift	0.906687775	8.121469425
19236	*S. paradoxus* shift	0.989881765	6.821984459
31505	*S. paradoxus* shift	0.955855579	3.032267535
32259	*S. paradoxus* shift	0.998665437	4.546902844
6506	*S. paradoxus* shift	0.982054204	5.468542886
16310	*S. paradoxus* shift	0.99652632	2.487101867
447	*S. paradoxus* shift	0.994506418	5.252074336
6281	Ornstein-Uhlenbeck	0.882367142	3.410968446
71042	*S. paradoxus* shift	0.804318406	6.030946867
6378	*S. cerevisiae* shift	0.845112064	1.00E-04
7165	*S. paradoxus* shift	0.811091269	4.465389345
6810	Ornstein-Uhlenbeck	0.859937275	2.618523967
6812	*S. paradoxus* shift	0.898839416	4.312524185
8150	*S. paradoxus* shift	0.999962114	2.871955612
6417	*S. paradoxus* shift	0.925463092	5.339113187
6407	*S. paradoxus* shift	0.988260506	8.792447836
462	*S. paradoxus* shift	0.817627126	7.291083934

Data are as in Table 1 except that inferences were made from experimental measurements of expression in purebred yeast homozygotes.

Supplementary Figure Legends



Supplementary Figure 1. Inferred values of parameters in simulations under an Ornstein-Uhlenbeck model of pathway regulatory evolution. In the main plot, each data point reports the results of one simulation of pathway regulatory evolution under an Ornstein-Uhlenbeck (OU) model in which the rates of regulatory evolution of pathway genes were drawn from an inverse-gamma distribution with $\alpha = 3$ and $\beta = 2$ and the OU constraint parameter θ was set to 10, after which parameter values for an OU model were optimized against the simulated expression data. For histograms at top and left, the independent variable is shared with the axis of the main plot and reports the indicated parameter value, and the dependent variable reports the proportion of simulated data sets in which the corresponding value was inferred. Note that inferences from most simulated data sets accurately estimate β and θ, but for a few data sets, large parameter values are inferred.

Supplementary Table Legends

Supplementary Table 1. Strains used in this work.

Supplementary Table 2. Fitted models of *cis*-regulatory evolution in yeast pathways. Data are as in Table 1 of the main text except that results for all pathways are shown.

Supplementary Table 3. Fitted models of species regulatory evolution in yeast pathways. Data are as in Table 2 of the main text except that results for all pathways are shown.